UNIVERSITA' DEGLI STUDI DEL MOLISE

DOTTORATO IN MEDICINA TRASLAZIONALE E CLINICA

XXXI CICLO

SETTORE SCIENTIFICO DISCIPLINARE BIO/16

"Meta-Analysis of DNA Methylation Profiles in Inflammatory Bowel Disease (IBD)"

Tutor

Prof. Germano Guerra

Candidata

Iolanda Agliata

Matr.155946

Coordinatore del Dottorato

Prof. Ciro Costagliola

Anno Accademico 2018/2019

Meta-Analysis of DNA Methylation Profiles in Inflammatory Bowel Disease (IBD)

1. Introduction

1.1- IBD: Crohn's disease and ulcerative colitis

Inflammatory bowel disease (IBD)-Crohn's disease (CD) and Ulcerative Colitis (UC)- are chronic and progressive inflammatory conditions of the gastrointestinal (GI) tract that affect about 2.2 million people in Europe and about 1.4 million in United States [1,2]. The exact etiology is not known, but IBD is characterized by various genetic abnormalities that result in aggressive response from both innate immunity, i.e. macrophages and neutrophils, and from acquired immunity, i.e. T cells and B cells [3]. In CD the intestinal portion affected by inflammation is the ileum, though the inflammation may affect the entire GI tract [4], while UC is characterized by chronic and relapsing inflammation of the colon and rectum [5] and is associated with increased risk of colon cancer development [6].

While genetics explains a fraction of inheritance of IBD (13,1% variance in CD and 8,2% in UC), environmental factors are able to influence susceptibility through epigenetic alterations, such as DNA methylation[7] . Epigenetics refers to heritable changes in phenotype that are independent of changes in the DNA sequence [8]. The association of differentially methylated DNA within regulatory regions of genes (promoters or sites of initiation of transcription) and gene repression has great biological significance[9]. However, the mechanism by which DNA methylation

regulates transcription is yet to be clarified. While methylated CpGs can prevent the binding of some transcription factors, whether gene silencing precedes or follows methylation is still debated. Moreover, several studies indicate that the effect on gene transcription is highly influenced by the location on the genome of the DNA methylated sequence [10].

There is also a known crosstalk between DNA methylation and inflammatory signals. Some studies have correlated the exposure of intestinal epithelial cells (IECs) to the proinflammatory cytokine IL-6 with the stabilization of the DNA methylation enzyme DNMT1 resulting in an increased methylation of DNA [11]. It is unclear whether the altered DNA methylation of epithelial cells could be due to persistent exposure of the colonic epithelium to inflammation in UC. Another cytokine that affects susceptibility to IBD is TGF-beta. It has been shown that TGF-beta germline null mice develop extensive inflammatory lesions in the colon at just few weeks of age[12,13]. In fact, damaged TGF-beta signaling in patients with IBD is associated with a reduced differentiation of Treg cells[14]. In both, immune and epithelial cells, TGF-beta has been shown to induce DNA methylation changes.[15]

2. Focus

Currently, the most used biomarkers for IBD are C-Reactive Protein and Calprotectin, although they are not specific for inflammation of intestinal origin, limiting their clinical use[16]. However, as DNA methylation is known to be tissue specific, it may represent a sensor of cytokine exposures and a better biomarker of IBD. Moreover, DNA markers are advantageous in terms of stability, improved isolation and storage and DNA methylation profiles represent a chemically more stable source for molecular diagnosis than RNA or many proteins[17]. With these assumptions, we performed a meta-analysis of intestinal epithelium methylomes in IBD. Our goal was to identify candidate “methyl-sensitive” loci that can be potentially useful as biomarkers, using base-resolution methylation data in mucosal biopsies from a large aggregated dataset of CD and UC patients, an approach that may open the way to personalized medicine.

3.Methods

3.1 - Dataset Selection

Dataset selection criteria included: methylome data obtained from intestinal mucosa, availability of healthy controls and IBD samples (CD, UC, or both), in data obtained using Human Infinium Bead Arrays (Illumina's HM450 or EPIC arrays), an established technology to detect DNA methylation [18]. Table 1 shows the main characteristics of the datasets fulfilling these criteria.

3.2- Data Preprocessing

All methylation data and sample information were downloaded from Gene Expression Omnibus (GEO) and Array Express public repositories, and analysed using R/Bioconductor packages [19]. Normalized data was loaded into R directly from each repository, except when raw idat files were also available. In that case, idat files were normalized using the "Funnorm" function of the minfi package [20]. Each dataset was independently assessed for data quality and distribution, before merging. Merged data was filtered for sex chromosomes, known cross-reactive probes [21], and probes associated with common SNPs that may reflect underlying polymorphisms rather than methylation profiles [22]. In addition, the "nmode.mc" function of the ENmIx package was used for the identification

of multimodal sites [23]. These sites were not removed at this step, but were used instead to classify significant associations in a later step.

3.3- Quality Control and Batch Correction

After filtering, 393112 CpG sites common to all datasets were used to identify principal components (PC) of variation and plotted using PC regression and multidimensional scaling (MDS) plots. Strong associations were observed between PCs and known variables (i.e. dataset, sex, age, and anatomical location), with age and anatomical location partially confounded by the dataset of origin. Latent variables were also identified, using surrogate variable analysis [24]. As additional quality control, DNA methylation values were used to predict age and sex and contrast with downloaded phenotype information. Sex was inferred from the median total intensity signal on XY chromosomes, and permitted the identification of 8 sex mismatches that were removed from the analysis (Fig S1). Age prediction was performed using Horvath's coefficients [25], as implemented in the wateRmelon package[26] . There was an overall correlation between reported and predicted age (Fig S1). For two datasets where age was not available, predicted age corresponded to adult samples, as reported in the corresponding repositories. The common merged and filtered matrix of methylation beta values and their corresponding phenotype data was taken to the next step.

3.4- Differential Methylation

Associations were tested for 393112 CpG sites, across 285 samples (81 control and 204 IBD samples). Methylation data was modeled at the probe and region levels using a linear model with Bayesian adjustment[27] . Sex and dataset were modeled together with subject status (i.e. control or IBD patient). Surrogate variables identified in the previous step were also included in the linear model to account for unknown sources of variation. Quantil-quantile (QQ) plots were used to inspect the distribution of resulting p values and estimate statistical inflation. Differentially methylated positions (DMPs) and regions (DMRs) were selected based on a methylation change (delta beta) of at least 10% when comparing control vs. IBD samples, and a false discovery rate- (FDR) adjusted p value below 0.05. DMRs were identified with the DMRcate package using the recommended proximity-based criteria[28] . A DMR was defined by the presence of at least two differentially methylated CpG sites with a maximum gap of 1000 bp. To identify CpG positions exhibiting significant differential variation and differential methylation (DVMCs), data was analysed using iEVORA, an algorithm that identifies DNA methylation outlier events shown to be indicative of malignancy[29] . iEVORA is based on Bartlett's test (BT) that examines the differential variance in DNA methylation, but because BT is very sensitive to single outliers, it is complemented

with re-ranking of significant events according to t-statistic (TT, t test), to balance the procedure. The significance is thus assessed at the level of differential variability, but the significance of differential variability with larger changes in the average DNA methylation are favored over those with smaller shifts. We used adjusted q(BT) <0.001 and p(TT) <0.05 as thresholds for significant DVMCs. To study genomic context, we used HM450 annotations, with hg19 as the human reference genome, UCSC and previously reported genomic features [30]. Differentially methylated genes (DMPs, DMRs, and DVMCs) were further analyzed to determine functional pathways and ontology enrichment using Enrichr [22].

3.5- SNPs-DMPs associations

To identify methylation quantitative trait loci (mQTL), single nucleotide polymorphisms (SNPs) associated with IBD risk were obtained from three independent studies: 1. Jostins L et al.[31] . 2. Huang H et al.[32] , and 3. Lange KM de et al.[33] . The genomic distances between 368 unique SNPs pooled from these three studies and IBD-associated DMPs were calculated using the R package GenomicRanges.

4.Results

4.1- Genome-wide changes in DNA methylation are a common feature of IBD

To identify DNA methylation changes in intestinal mucosa associated with IBD, we reanalyzed bead-array methylation data from different datasets (**Table 1**). Samples from these datasets included pediatric and adult IBD patients, from both sexes, and involved the two main forms of the condition (i.e. CD and UC). After filtering (see Methods), we tested for the association between IBD and DNA methylation at 393112 CpG sites (81 control and 204 IBD patients) using a linear model. In such a model, we adjusted for sex, age, dataset, and surrogate variables identified during data preprocessing (Fig S1). To account for statistical inflation, we used criteria of effect size (change in mean methylation of at least 10% between controls and IBD) and FDR-adjusted p value < 0.05.

Using these criteria, we identified 4280 differentially methylated positions (DMPs), out of which 437 were hypo and 3843 were hypermethylated in IBD (Fig 1A, Table 2 and Table S1). DMPs were robust to IBD type (Fig 1B), other clinical features (Fig 1C) and study of origin (Fig S1). Although many of these sites were previously identified, in particular in the larger dataset published by Howell et al[34], new associations were identified in our reanalysis of the combined datasets.

4.2- Genomic context of IBD-associated DNA methylation

DMPs distinguishing IBD from control tissues were assessed for genomic distribution, in terms of gene-centric and CpG island (CGI)-centric context. DMPs were relatively absent from CGIs, gene promoters, or the vicinity of transcription start sites (TSS) (Fig 2A-2C). Instead, hypo and hypermethylated DMPs were highly concentrated in non-CGI regions (i.e. open sea) (Fig 2A). In addition, an important proportion of DMPs were close to each other, suggesting a non-random association with particular genomic loci. To explore this observation, we performed region-level analysis in the same combined dataset. This led to the identification of 1017 differentially methylated regions (DMRs), 172 hypo and 845 hyper methylated in IBD (Table 2 and S2). As expected, many of these regions corresponded to gene loci also identified using the probe-level strategy (Fig S2).

4.3- Inflammation-related pathways are enriched in DNA methylation changes associated with IBD

Pathway analysis of DMRs showed over-representation of pathways related to metabolism and signal transduction, such as: Adipogenesis genes, Hemostasis, G alpha signaling events, Pathways in cancer, TGF-beta Receptor Signaling (Table 4).

Finally, rather than mean methylation values, methylation variation has been associated with disease and cancer susceptibility [35]. To explore this, we used the iEVORA algorithm in the same datasets, to identify differentially variable and methylated CpGs (DVMCs). Using stringent criteria of differential methylation and variation, we identified more than 15K DVMCs (Fig 2D and Table S3). Of note, for most of these sites, IBD samples displayed higher variability than control tissues.

4.4- SNPs associated with IBD risk are genomically closer to hypomethylated DMPs

DNA methylation may represent an intermediary between genotype and disease susceptibility. Of note, an important fraction of DMPs fall within susceptibility loci previously identified in GWAS (Table 1 and S1). Among DMRs with a significant genetic association, confirming the findings of previous studies, we find ITGB2,

MUC16, JAK3, KRT8, HLA genes. Moreover, some DMPs display a bimodal DNA methylation distribution, suggesting that their methylation levels are directly dependent on genotype. To explore a genotype-methylation association, we calculated the distance between DMPs identified in our analysis and SNPs associated with IBD risk (pooled from three different studies)[31–33] . Of note, hypomethylated DMPs were overall significantly closer to a known IBD risk SNP, compared to hypermethylated or control DMPs, and this difference was consistent across the three independent SNP datasets (Fig 2C and S2C).

5.Conclusions

IBDs are complex pathologies that present a wide range of phenotypes and different trajectories. It is therefore necessary to establish novel molecular signatures that are efficient in diagnostics and that allow a better distinction of clinical cases that require a first timely treatment from those who have a quiescent course of the disease. The importance of using molecular signatures in IBD is partly due to their stability. A high consistency in IBD-associated methylation signatures at different time points from the same patient, has been described[34]. This suggested that the stability of epigenetic alterations could contribute to chronic recurrent inflammation via an altered IEC function[34]. Among the DMRs with a significant genetic association, confirming the findings of previous studies, we find *ITGB2, MUC16, JAK3, KRT8, HLA* genes. ITGB2 has a role in leukocyte adhesion and activation, which is particularly interesting since the recent focus of many therapeutic strategies is leukocyte adhesion, as well as that of vedolizumab[36]. MUC16 in previous studies was found to be significantly increased in the active form of UC compared to the control group both in the mucosa and in the submucosa, neverthless no association between the expression of MUC16 and clinical characteristics of patients with UC has been found [37]. KRT8 (Keratin-8) variants could play a role in sporadic forms of IBD, as studies have shown that KRT8-null mice developed spontaneous colitis and liver damage[38]. From several studies it is already known how the inhibition of JAK3, a family kinase member, by tofacitinib involves a considerable reduction of inflammation by

interfering with different cytokine receptors[39]. Our study proposes further details about enrichment in disease-specific DMRs and DMPs that may be a useful diagnostic tool. Several studies have highlighted the linkage between the IBD3 region (6p21.1-23), a region which includes the human leukocyte antigen (HLA) complex and CD or UC. Some studies have calculated the risk of sharing the HLA allele within families and it is emerged that this region is a genetic risk factor in CD for 10%-33% and for 64%-100% in UC [40]. Nevertheless, this region remains an unclear risk factor due to the extensive polymorphisms that characterize it, to the high gene density and the high variability of allele frequencies in different populations[40].

The presence of hypomethylated sites in the vicinity of known susceptibility loci supports the notion of DNA methylation as an intermediary between genotype and phenotype (mQTLs). Although it is not clear why such SNP-DMP association was only observed with hypomethylated DMPs, we ruled out a potential global hypomethylation in IBD (data not shown). A possible explanation of this phenomenon derives from a study that explained the correlation between genome-wide genetic variation and proximal DNA methylation patterns[41]. In fact, 5-methyl-cytosines at CpG sites often change into thymines, which involves a large number of spontaneous mutations. Often the repair system could commit a significant number of errors in neighboring regions if the synthesis of erased gaps around deaminated 5-methyl-cytosines is error-prone. 18% of the human genome is located within 10

bp of CpG sites , in these regions the SNP rate increases by about 50% if the neighboring CpG sites are methylated. In hypomethylated regions, the CGCG motif is enriched and evolutionary conserved and it seems that there is a slow CpG deamination rather than fast CpG gain[41]. This means that the CGCG motif is a possible cis-element to mantain the hypomethylated state. Therefore, our hypothesis is that there is likely to be a positive selective pressure that causes SNPs to be close to hypomethylated regions so that the mutational rate does not increase. However, most of the changes that we described in IBD are instead due to hypermethylation. In addition, our study suggest a possible interaction between the HLA-genes showed in our panel and DNA methylation in patients carrying these variants. Furthermore, our study shows for the first time an enrichment in methylation changes in CpG-poor regions (i.e. open sea) in IBD, suggesting that methylation of these non-coding regions may be relatively dynamic in certain conditions.

The pathway analysis of DMRs shows an over-representation of pathways of the metabolism and signal transduction. This analysis reveals some differences with the highlighted pathways analysis emerged from the work of Howell et al., which collects the largest group of samples reviewed by us. Howell's pathways analysis was enriched in immune system pathways suggesting an important role of the immune system in the defense of the host and that alterations in these processes could lead to the beginning of intestinal inflammation in the two conditions (UC and

CD). From our data it emerges that the most dysregulated pathways are those related to metabolism and signal transduction. In particular, the TGF-beta pathway plays a crucial role in regulating the intestinal immune response in particular of T cells, through a molecular mechanism recently demonstrated by Fenton TM[42] and his group. A TGF-beta activator is the integrin αvβ8 whose expression is strongly upregulated in intestinal dendritic cells of IBD patients[42].

Another fundamental element that emerged from our analysis is the crosstalk between inflammatory bowel diseases and adipogenesis. In fact, patients with IBD, particularly those with CD, have fat-wrapping or creeping-fat that corresponds to the ectopic adipose tissue extending from the mesenteric area and cover the surface of a large part of the small and large intestine. Inflammation plays a role, with mesenteric adipose tissue of patients with IBD presenting numerous morphological and functional alterations with infiltrates of immune cells such as T cells and macrophages. It is likely that in obese (BMI>30) or overweight IBD patients (BMI=25-30) it is precisely the mesenteric adipose tissue that contributes to intestinal and systemic inflammation[43]. As a limitation to our study, body mass index data of the same patients profiled for DNA methylation was not available, so this hypothesis remains to be verified.

Finally, in light of the fact that methylation variation has been associated with the risk to develop cancer our analysis is also enriched in DVMC. The iEVORA algorhitm

is able to identify differentially variable and differentially methylated CpGs with the characteristic that become more variable and hypermethylated in normal samples near a tumoral tissue. In these hypervariable and hypermethylated regions next to a tumor tissue there are changes in DNA methylation beta values of the order of 20%-30% [35].

In summary, our findings illustrate an aberrant DNA methylation landscape in IBD, that is independent of IBD subtype and other clinical and pathological features. Following pospective studies, such landscape may be a useful source of biomarkers of IBD risk and progression.

Bibliography

1. Loftus, E. V. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* **126**, 1504–1517 (2004).

2. Saleh, M. & Elson, C. O. Experimental inflammatory bowel disease: insights into the host-microbiota dialog. *Immunity* **34**, 293–302 (2011).

3. Legaki, E. & Gazouli, M. Influence of environmental factors in the development of inflammatory bowel diseases. *World J Gastrointest Pharmacol Ther* **7**, 112–125 (2016).

4. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).

5. Head, K. A. & Jurenka, J. S. Inflammatory bowel disease Part 1: ulcerative colitis--pathophysiology and conventional and alternative treatment options. *Altern Med Rev* **8**, 247–283 (2003).

6. Herrinton, L. J. *et al.* Incidence and Mortality of Colorectal Adenocarcinoma in Persons With Inflammatory Bowel Disease From 1998 to 2010. *Gastroenterology* **143**, 382–389 (2012).

7. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979–986 (2015).

8. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev* **23**, 781–783 (2009).

9. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**, 484–492 (2012).

10. Medvedeva, Y. A. *et al.* Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics* **15**, 119 (2014).

11. Foran, E. *et al.* Upregulation of DNA Methyltransferase–Mediated Gene Silencing, Anchorage-Independent Growth, and Migration of Colon Cancer Cells by Interleukin-6. *Mol Cancer Res* **8**, 471–481 (2010).

12. Shull, M. M. *et al.* Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature* **359**, 693–699 (1992).

13. Kulkarni, A. B. & Karlsson, S. Transforming growth factor-beta 1 knockout mice. A mutation in one cytokine gene causes a dramatic inflammatory disease. *Am. J. Pathol.* **143**, 3–9 (1993).

14. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446–450 (2013).

15. Martin, M. *et al.* Dynamic imbalance between cancer cell subpopulations induced by transforming growth factor beta (TGF-β) is associated with a DNA methylome switch. *BMC Genomics* **15**, 435 (2014).

16. Soubières, A. A. & Poullis, A. Emerging role of novel biomarkers in the diagnosis of inflammatory bowel disease. *World J Gastrointest Pharmacol Ther* **7**, 41–50 (2016).

17. Laird, P. W. Early detection: The power and the promise of DNA methylation markers. *Nature Reviews Cancer* **3**, 253–266 (2003).

18. Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692–702 (2011).

19. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80 (2004).

20. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

21. Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, (2016).

22. Chen, Y. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8**, 203–209 (2013).

23. Xu, Z., Niu, L., Li, L. & Taylor, J. A. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res* **44**, e20–e20 (2016).

24. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

25. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115 (2013).

26. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

27. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).

28. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin* **8**, 6 (2015).

29. Teschendorff, A. E. *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* **7**, 10478 (2016).

30. Slieker, R. C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6**, 26 (2013).

31. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).

32. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).

33. Lange, K. M. de *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *bioRxiv* 058255 (2016). doi:10.1101/058255

34. Howell, K. J. *et al.* DNA Methylation and Transcription Patterns in Intestinal Epithelial Cells From Pediatric Patients With Inflammatory Bowel Diseases Differentiate Disease Subtypes and Associate With Outcome. *Gastroenterology* **154**, 585–598 (2018).

35. Teschendorff, A. E. *et al.* DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun* **7**, 10478 (2016).

36. Ventham, N. T. *et al.* Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun* **7**, 13507 (2016).

37. Yamamoto-Furusho, J. K., Ascaño-Gutiérrez, I., Furuzawa-Carballeda, J. & Fonseca-Camarillo, G. Differential Expression of MUC12, MUC16, and MUC20 in Patients with Active and Remission Ulcerative Colitis. *Mediators Inflamm.* **2015**, 659018 (2015).

38. Tao, G.-Z. *et al.* Analysis of keratin polypeptides 8 and 19 variants in inflammatory bowel disease. *Clin. Gastroenterol. Hepatol.* **5**, 857–864 (2007).

39. Vuitton, L., Koch, S. & Peyrin-Biroulet, L. Janus kinase inhibition with tofacitinib: changing the face of inflammatory bowel disease treatment. *Curr Drug Targets* **14**, 1385–1391 (2013).

40. Muro, M., López-Hernández, R. & Mrowiec, A. Immunogenetic biomarkers in inflammatory bowel diseases: role of the IBD3 region. *World J. Gastroenterol.* **20**, 15037–15048 (2014).

41. Qu, W. *et al.* Genome-wide genetic variations are highly correlated with proximal DNA methylation patterns. *Genome Res.* **22**, 1419–1425 (2012).

42. Fenton, T. M. *et al.* Inflammatory cues enhance TGFβ activation by distinct subsets of human intestinal dendritic cells via integrin αvβ8. *Mucosal Immunol* **10**, 624–634 (2017).

43. Gonçalves, P., Magro, F. & Martel, F. Metabolic inflammation in inflammatory bowel disease: crosstalk between adipose tissue and bowel. *Inflamm. Bowel Dis.* **21**, 453–467 (2015).

Figure Legends

Figure 1. DNA methylation distinguishes IBD from healthy tissues. A. Top differentially methylated positions (DMPs) with a mean difference between IBD (red) vs Control (gray) of at least 20% (delta-beta>20). Probe ID and corresponding nearest gene are shown for each significant CpG site. Methylation is represented on the y axis as normalized beta values. **B.** The same CpG sites shown in (A) are represented separately for ulcerative colitis (UC) and Chron's disease (CD), shown in blue and green, respectively. **C.** Heatmap showing top differentially methylated positions between IBD *vs* control. The red to blue color gradient represents higher to lower methylation. The main covariates considered in the analysis (i.e. dataset, localization, and sex) are also represented.

Figure 2. Genomic distribution of IBD-related DMPs. DMPs were annotated according to CpG islands (CGI) (**A)**, relation to gene features **(B)**, and distance to the nearest transcription start site (TSS) **(C).** For each genomic context, distribution is shown separately for all DMPs, those hypo or hypermethylated in IBD relative to healthy tissues, and all the HM450 probes, as a control. **2D** Top differentially variable methylated CpG sites (DVMCs) in IBD *vs* Control. DNA methylation was

plotted as beta values for each of the top nine DVMC identified with the iEVORA algorithm (see Methods section).

Supplementary Data

Figure S1. Data quality and preprocessing.

Figure S2. Supplementary results.

Table S1. Full list of differentially methylated positions (DMPs).

Table S2. Full list of differentially methylated regions (DMRs).

Table S3. Full list of differentially variable and methylated CpGs (DVMCs).

Table1.

Characteristics of the datasets included in the study.

Accession	Condition	Technique	idat	Samples	Age	Country	PMID
MTAB_5463	IBD	HM450/EPIC	yes	111/104	6-15	Europe	29031501
GSE32146	IBD	HM450	no	25	14-17	USA	30232239
MTAB_3703/3709	IBD	HM450	yes	12	14	Europe	26376367
GSE81211	UC	HM450	yes	12	unknown	South Korea	NA
GSE105798	CD	HM450	yes	11	unknown	South Korea	25228829
GSE42921	IBD	HM450	no	23	9-16	USA	NA

Table 2. Top DMPs.

Differentially methylated positions (DMPs) with a mean difference between groups of at least 20% (delta-beta > 20). Probe ID: Illumina probe reference, logFC: logarithmic fold-change between groups (IBD vs control), FDR: false discovery rate, Symbol: gene symbol, Distance: distance in base pairs to the closest gene. Full list of DMPs can be found in Supplementary Table S1.

Probe ID	logFC	FDR	Symbol	Distance
cg16465027	-1.14	3.5E-16	*PHACTR1*	122016
cg07839457	-1.48	1.3E-14	*NLRC5*	435
cg16240683	1.15	1.3E-14	*ZNF436-AS1*	0
cg19269426	-1.26	3.6E-14	*GGPS1*	0
cg22718139	1.36	1.2E-13	*HMGCS2*	0
cg26974214	-1.19	1.0E-12	*LIPA*	0
cg24129356	-1.33	1.1E-12	*HLA-DMA*	0
cg02806715	-1.17	9.6E-12	*HLA-DMA*	0
cg09321817	-1.63	1.4E-11	*HLA-DPA1*	0
cg01804934	-1.28	1.4E-10	*HLA-DPA1*	0
cg23045908	-1.43	1.0E-	*PDE4B*	0

cg06061086	-1.05	09 1.1E-09	*FOXP4*	14522

Table 3. Top DMRs.

Differentially methylated regions (DMRs) with at least three CpG sites, and a maximum beta change between groups of at least 10%. # CpGs: number of CpG sites per region, FDR: false discovery rate, Beta FC: methylation beta value fold change. Full list of DMRs can be found in Supplementary Table S2.

Symbol	Coordinates	# CpG	FDR	Beta FC
HOXA6	chr7:27180888-27185512	44	0.0E+00	0.11
HOXA3	chr7:27152583-27156062	28	1.2E-234	0.11
HOXA3	chr7:27159883-27166103	18	2.0E-79	0.11
HNF4A	chr20:42983920-42984878	12	6.2E-140	0.10
HLA-DPB1	chr6:33046344-33049505	22	5.9E-204	0.17
FABP1	chr2:88427027-88428542	8	5.9E-97	0.10
HNF1A	chr12:121415506-121416796	8	3.4E-105	0.10
DENND1C	chr19:6475456-6477198	7	3.2E-68	0.11
LINC00982	chr1:2979311-2980937	8	2.3E-88	0.10
DUSP6	chr12:89747628-89749822	15	1.1E-108	0.11
HLA-DPA1	chr6:33040535-33041697	8	6.5E-87	-0.22
ELMO3	chr16:67231928-	9	6.2E-89	0.11

Symbol	Coordinates	# CpG	FDR	Beta FC
	67233983			
HNF4A	chr20:43028501-43029997	9	4.6E-93	0.11
PRR26	chr10:695301-696356	10	2.3E-88	0.10
QTRT1	chr19:10822935-10824846	7	7.7E-75	0.10
PLGLA	chr2:106959205-106959878	7	2.2E-75	0.11
BST2	chr19:17516282-17517008	5	2.2E-83	-0.15
HOXA-AS3	chr7:27178861-27179432	5	2.1E-80	0.11
FMNL1	chr17:43318045-43319382	7	2.0E-75	0.11
LAMB3	chr1:209825672-209825856	6	1.2E-76	0.11
LAMA3	chr18:21452730-21452895	6	8.0E-74	0.10
PSMG3	chr7:1606266-1607787	8	6.3E-71	0.12
GATA6	chr18:19756582-19758221	7	9.0E-64	0.11
SPPL2B	chr19:2278451-2278847	5	3.4E-62	0.10
RASAL3	chr19:15568360-15568935	5	7.6E-62	0.10
LGALS3	chr14:55602634-55604454	4	1.3E-48	0.13
TNNC1	chr3:52487733-52488229	5	6.1E-74	0.12
HMGCS2	chr1:120311439-120311653	4	1.8E-57	0.20
SPIRE2	chr16:89913618-89913972	4	1.9E-55	0.12
PLEKHM3	chr2:208794914-208795859	5	2.1E-49	0.11
RABGAP1L	chr1:174843754-174844490	5	3.6E-51	0.11
MIR3193	chr20:30195969-30196714	5	7.3E-53	0.11

Symbol	Coordinates	# CpG	FDR	Beta FC
FCMR	chr1:207095153-207096833	6	3.1E-45	-0.11
IGF2BP1	chr17:47090616-47092178	8	4.5E-59	0.11
MUC13	chr3:124652790-124653718	4	3.6E-36	0.10
VILL	chr3:38033516-38033934	3	8.5E-48	0.14
VWA1	chr1:1368846-1370775	9	9.1E-67	-0.18
THBS1	chr15:39874776-39876248	4	9.1E-31	0.12
RNU5F-1	chr1:220132091-220132728	6	5.4E-53	0.11
CIITA	chr16:10969805-10971250	6	2.2E-54	-0.13
SEC16B	chr1:177939225-177939546	4	2.5E-43	0.10
SH2D3C	chr9:130515119-130517848	6	2.7E-43	0.11
EPHB3	chr3:184297380-184297522	3	4.4E-43	-0.18
TMED6	chr16:69385547-69386847	6	7.5E-58	0.10
ADGRG1	chr16:57653169-57654347	5	2.8E-58	0.11
KRT8	chr12:53320306-53321588	3	5.5E-31	0.10
LGALS4	chr19:39303506-39305100	5	4.0E-38	0.10
DACT2	chr6:168665386-168665533	3	4.0E-42	0.10
BTNL8	chr5:180325254-180326186	5	1.2E-48	0.10
GAS2	chr11:22696087-22696308	3	1.0E-38	0.11
RNF19A	chr8:101348456-101348501	3	2.9E-38	0.10
PNLIPRP2	chr10:118380175-118381044	3	1.1E-32	0.11
C12orf75	chr12:105956397-	3	4.3E-38	0.11

Symbol	Coordinates	# CpG	FDR	Beta FC
	105956478			
LINC01091	chr4:124468530-124468840	3	3.9E-36	0.10
COMTD1	chr10:77021847-77021880	3	2.5E-39	0.12
MEP1A	chr6:46761035-46761090	3	5.5E-36	0.10
TOLLIP	chr11:1325718-1327450	10	3.2E-52	0.12
DAPP1	chr4:100737138-100738139	5	1.3E-29	-0.11
EPHB3	chr3:184290594-184290794	4	4.1E-34	-0.12
ZNF436	chr1:23696021-23698143	8	5.8E-60	0.17
DOT1L	chr19:2213373-2214108	4	3.0E-32	0.11
GPR160	chr3:169758289-169759041	3	2.9E-36	0.11
KIDINS220	chr2:8849962-8850066	3	1.0E-32	0.10
JAK3	chr19:17942221-17943313	5	3.4E-47	0.10
FAAP20	chr1:2120985-2121724	6	5.0E-34	0.11
APOH	chr17:64225346-64226953	5	2.3E-34	0.11
PDZK1	chr1:145726979-145727762	6	6.9E-54	0.11
HLA-B	chr6:31322298-31322926	4	1.3E-30	-0.11
TERT	chr5:1291888-1293231	4	6.3E-39	0.10
ZFHX3	chr16:73090646-73091565	4	2.3E-39	0.11
KLB	chr4:39407527-39408665	4	3.9E-38	0.10
HNMT	chr2:138721315-138721836	3	1.3E-32	0.11
MACROD1	chr11:63852804-63853037	5	1.8E-28	0.10

Symbol	Coordinates	# CpG	FDR	Beta FC
MAP3K5	chr6:136914796-136915556	4	5.2E-26	-0.14
RARG	chr12:53612551-53613154	4	6.0E-45	0.10
AATK	chr17:79107083-79108224	7	1.6E-37	0.10
HLX	chr1:221057236-221059414	8	2.5E-22	0.13
LINC02372	chr12:127544951-127545433	6	1.8E-32	0.10
BTBD9	chr6:38145125-38145464	3	3.4E-25	0.11
SMAD3	chr15:67356838-67357361	3	2.9E-33	0.10
LPP	chr3:187930521-187930704	3	3.7E-26	0.10
EIF2AK4	chr15:40268421-40269214	5	8.0E-24	0.11
AFF3	chr2:100170766-100171136	4	1.3E-21	0.13
SH3BP2	chr4:2813458-2814122	5	2.6E-23	-0.10
TACC3	chr4:1762460-1763211	3	2.8E-19	0.10
PTGDR2	chr11:60619955-60621110	3	3.6E-37	0.12
PARP9	chr3:122281881-122281975	3	4.2E-22	-0.12
SFT2D3	chr2:128453108-128453484	5	9.0E-19	0.11
TTBK1	chr6:43244304-43245571	3	1.4E-17	0.10
BMP4	chr14:54418728-54420185	6	9.8E-21	0.11
ITGB6	chr2:161056600-161057625	3	3.6E-31	0.13
IL12RB1	chr19:18197544-18198611	5	1.3E-32	-0.13
TRIM69	chr15:45018591-45018905	3	5.0E-28	-0.13
HLA-DRA	chr6:32407289-	8	4.2E-24	-0.13

Symbol	Coordinates	# CpG	FDR	Beta FC
HLA-C	32408284 chr6:31238245-31238751	3	1.9E-21	-0.11
CACNA1A	chr19:13318873-13319777	4	1.4E-19	-0.13
ATXN7L1	chr7:105279391-105279882	3	9.5E-16	0.10
TRIO	chr5:14405632-14406585	5	4.5E-13	0.12
LPIN1	chr2:11917490-11917787	3	3.8E-11	-0.12
LIPA	chr10:91151885-91152447	3	9.5E-17	-0.17

Table 4. Pathway analysis with adjusted p-value <0.05

NAME	ADJ. P-VALUE	COMB.SCORE	Enrichr
Adipogenesis genes_Mus musculus_WP447	0.000005781	33.17	WikiPathways 2016
Adipogenesis genes_Homo sapiens_WP236	0.000005781	32.10	WikiPathways 2016
TGF Beta Signaling Pathway_Mus musculus_WP113	0.0004283	26.34	WikiPathways 2016
TGF-beta Receptor Signaling_Homo sapiens_WP560	0.0007316	24.45	WikiPathways 2016
Alpha6-Beta4 Integrin Signaling Pathway_Mus musculus WP488	0.002441	18.52	WikiPathways 2016
Hemostasis_Homo sapiens_R-HSA-109582	0.001314	29.15	Reactome
G alpha(12/13) signalling events_Homo sapiens_R-HSA-416482	0.004290	21.79	Reactome
Pathways in cancer _Homo	0.0005900	26.06	KEGG 2016

sapiens_hsa 05200 Aldosterone synthesis and secretion_Homo sapiens_hsa04925	0.0005900	24.68	KEGG 2016
Cholinergic synapse_Homo sapiens_hsa04725	0.004515	19.59	KEGG 2016
AGE-RAGE signaling pathway in diabetic complications_Homo sapiens_hsa04933	0.004515	18.97	KEGG 2016
Heterotrimeric G protein signal. pathway-Gq alpha and Go alpha med. Path-hsaP00027	0.005992	15.17	Panther 2016
Histamine H1 receptor mediated signaling pathway_Homo sapiens_P04385	0.005992	11.96	Panther 2016

UNIVERSITÀ DEGLI STUDI DEL MOLISE

DIPARTIMENTO DI MEDICINA E DI SCIENZE DELLA SALUTE "VINCENZO TIBERIO"

MEDAGLIONE DOTTORANDA XXXI CICLO
DR.SSA AGLIATA IOLANDA

TITOLO	DNA METHYLATION PROFILE IN A LARGE COHORT OF IBD PATIENTS
TUTOR	PROF. GERMANO GUERRA
PRODOTTI ATTIVITÁ DI RICERCA	ALLEGATO 1A.1
RELAZIONE SINTETICA	ALLEGATO 1A.2
PARTECIPAZIONE A CORSI DI INSEGNAMENTO	21-set-18 *Il fenotipo longevità: genetica o stile di vita*? Prof. Calogero Caruso 21-set-18 *Nutrigerontologia: una disciplina innovativa per contrastare il declino di funzione età correlato.* Dott. Sergio Davinelli.
GIUDIZIO COMPLESSIVO	Positivo

Il Coordinatore
Prof. Ciro Costagliola

UNIVERSITÀ DEGLI STUDI DEL MOLISE

Translational and Clinical Medicine
Evaluation of the PhD thesis

Candidate: **IOLANDA AGLIATA** **iolanda.aglinta87@gmail.com**

Tutor: **GERMANO GUERRA** **germano.guerra@unimol.it**

Title: **Meta-Analysis of DNA Methylation Profiles in Inflammatory Bowel Disease (IBD)**

Evaluator Professor: **HECTOR HERNANDEZ-VARGAS**

Please provide a quantitative evaluation, according to the following parameters
Excellent: 5, Good: 4, Fair: 3, Poor: 2, Very Poor: 1

	EVALUATION (from 1 to 5)
Does the thesis deal with relevant issues according to the subject areas of the PhD program (Translational and Clinical Medicine)?	5
Are the objectives of the thesis clearly stated?	4
Which is the level of originality of the elaborate?	4
Are the tools and methods used appropriate to the objectives to be achieved?	5
Are the results obtained adequately presented?	4
Are the figures clear and useful to better understand the text?	4
Are results and interpretations clearly distinct?	3
Are the general interpretations and conclusions adequately supported by the data presented?	4
Does the thesis present original data and / or conclusions?	4
Is the text clear?	4
Is the used terminology correct?	5
Is the bibliography adequate and correctly cited?	5
Are the captions of the figures correct and complete?	3

Any other comment / suggestion (use other pages if it is necessary):

The work developed by Iolanda AGLIATA involves a highly relevant subject in Translational Medicine: the identification of molecular biomarkers with potential clinical utility. While developing her writing and analytical skills, she contributed to a meta-analysis of genome-wide DNA methylation data that is expected to lead to a scientific publication.

Overall Evaluation (please indicate one option):

☑	The candidate can be admitted to the Final Exam
☐	The candidate can be admitted to the Final Exam, but the thesis requires a minor revision, without requiring further evaluation by the evaluator
☐	To be admitted to the final exam, the thesis requires an extensive review. The revised version of the thesis will have to be elaborated in maximum 6 months, and re-examined by the evaluator.

Place and date: Lyon 22/12/18 H. Hernandez *Signature*

Please re-send this evaluation to:

Translational and Clinical Medicine
Evaluation of the PhD thesis

Candidate: IOLANDA AGLIATA _____iolanda.agliata87@gmail.com_____

Tutor: GERMANO GUERRA_____germano.guerra@unimol.it_____

Title: *Meta-Analysis of DNA Methylation Profiles in Inflammatory Bowel Disease (IBD)*

Evaluator Professor: GIULIA DE FALCO – ___Queen Mary University of London___giulia.defalco@qmul.ac.uk___

Please provide a quantitative evaluation, according to the following parameters:

Excellent: 5
Good: 4
Fair: 3
Poor: 2
Very Poor: 1

	EVALUATION (from 1 to 5)
Does the thesis deal with relevant issues according to the subject areas of the PhD program (Translational and Clinical Medicine)?	5
Are the objectives of the thesis clearly stated?	4
Which is the level of originality of the elaborate?	4
Are the tools and methods used appropriate to the objectives to be achieved?	5
Are the results obtained adequately presented?	4
Are the figures clear and useful to better understand the text?	4
Are results and interpretations clearly distinct?	4
Are the general interpretations and conclusions adequately supported by the data presented?	5
Does the thesis present original data and / or conclusions?	5
Is the text clear?	4
Is the used terminology correct?	5
Is the bibliography adequate and correctly cited?	5
Are the captions of the figures correct and complete?	4

Any other comment / suggestion (use other pages if it is necessary):

Overall Evaluation (please indicate one option):

☒	The candidate can be admitted to the Final Exam
☐	The candidate can be admitted to the Final Exam, but the thesis requires a minor revision, without requiring further evaluation by the evaluator
☐	To be admitted to the final exam, the thesis requires an extensive review. The revised version of the thesis will have to be elaborated in maximum 6 months, and re-examined by the evaluator.

Place and date:
London 14-02-2019
_____

Signature
_____

1

UNIVERSITÀ DEGLI STUDI DEL MOLISE

Translational and Clinical Medicine
Evaluation of the PhD thesis

Candidate: IOLANDA AGLIATA **iolanda.agliata87@gmail.com**

Tutor: GERMANO GUERRA **germano.guerra@unimol.it**

Title: **Meta-Analysis of DNA Methylation Profiles in Inflammatory Bowel Disease (IBD)**

Evaluator Professor: JULIEN MARIE

Please provide a quantitative evaluation, according to the following parameters:

Excellent: 5
Good: 4
Fair: 3
Poor: 2
Very Poor: 1

	EVALUATION (from 1 to 5)
Does the thesis deal with relevant issues according to the subject areas of the PhD program (Translational and Clinical Medicine)?	5
Are the objectives of the thesis clearly stated?	4
Which is the level of originality of the elaborate?	4
Are the tools and methods used appropriate to the objectives to be achieved?	4
Are the results obtained adequately presented?	3.5
Are the figures clear and useful to better understand the text?	4
Are results and interpretations clearly distinct?	4
Are the general interpretations and conclusions adequately supported by the data presented?	3
Does the thesis present original data and / or conclusions?	4
Is the text clear?	4
Is the used terminology correct?	5
Is the bibliography adequate and correctly cited?	5
Are the captions of the figures correct and complete?	3

Any other comment / suggestion (use other pages if it is necessary):
Iolanda is an extremely motivate student. Her work developed in the lab allows to contribute to metagenomic analysis question we are addressing in IBD. In addition, Iolanda performed a deep literature research on the topic of her PhD project. Both the bench work and the literature work should give rise to a publication soon.

Overall Evaluation (please indicate one option):

x	The candidate can be admitted to the Final Exam
☐	The candidate can be admitted to the Final Exam, but the thesis requires a minor revision, without requiring further evaluation by the evaluator
☐	To be admitted to the final exam, the thesis requires an extensive revision. The revised version of the thesis will have to be elaborated in maximum 6 months, and re-evaluated by the evaluator

Place and date: Lyon 02/01/2019

CENTRE DE RECHERCHE EN CANCÉROLOGIE DE LYON (CRCL)
UMR INSERM 1052 - CNRS 5286
28 rue Laennec - Centre Léon Bérard
69373 Lyon Cedex 08

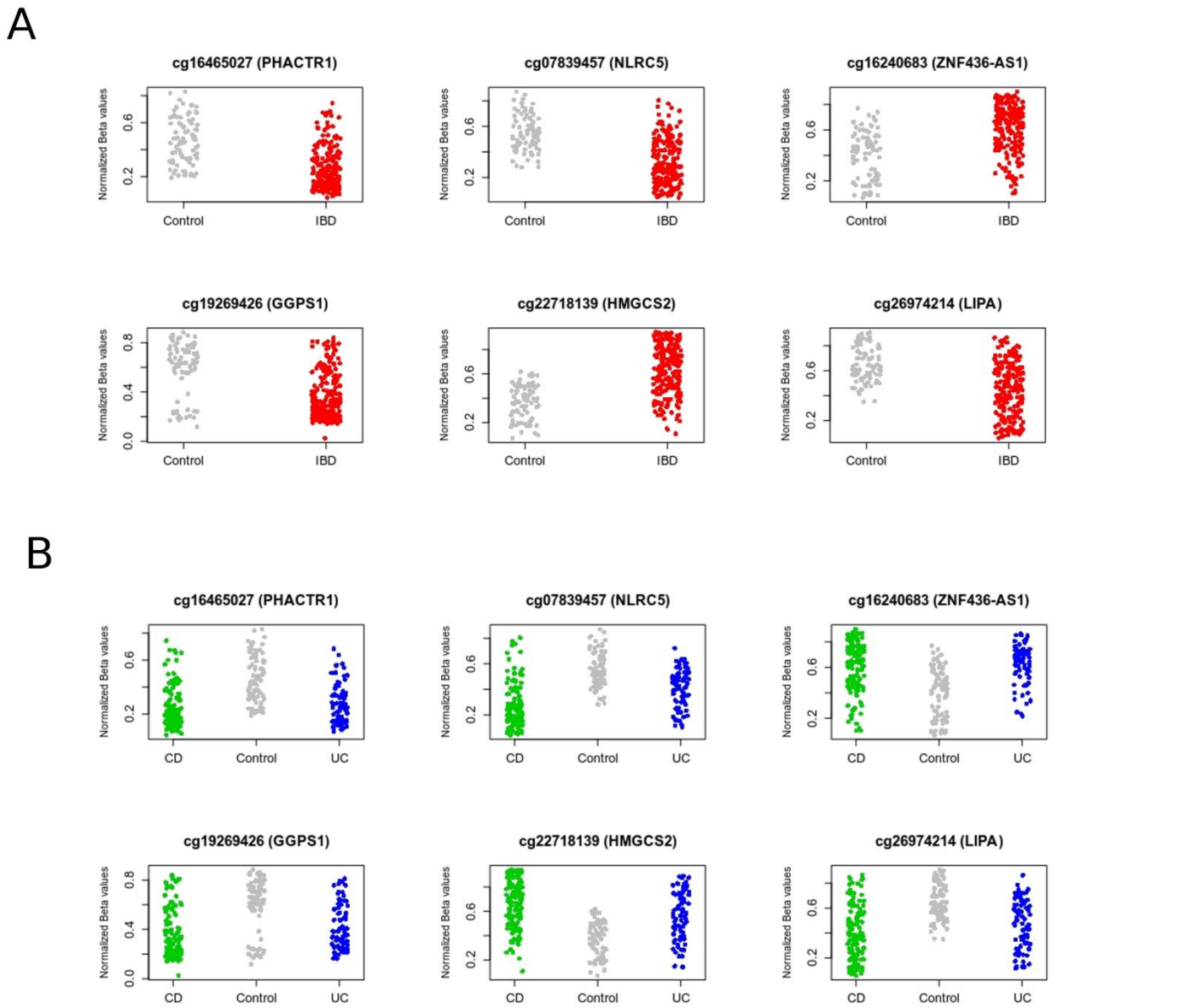
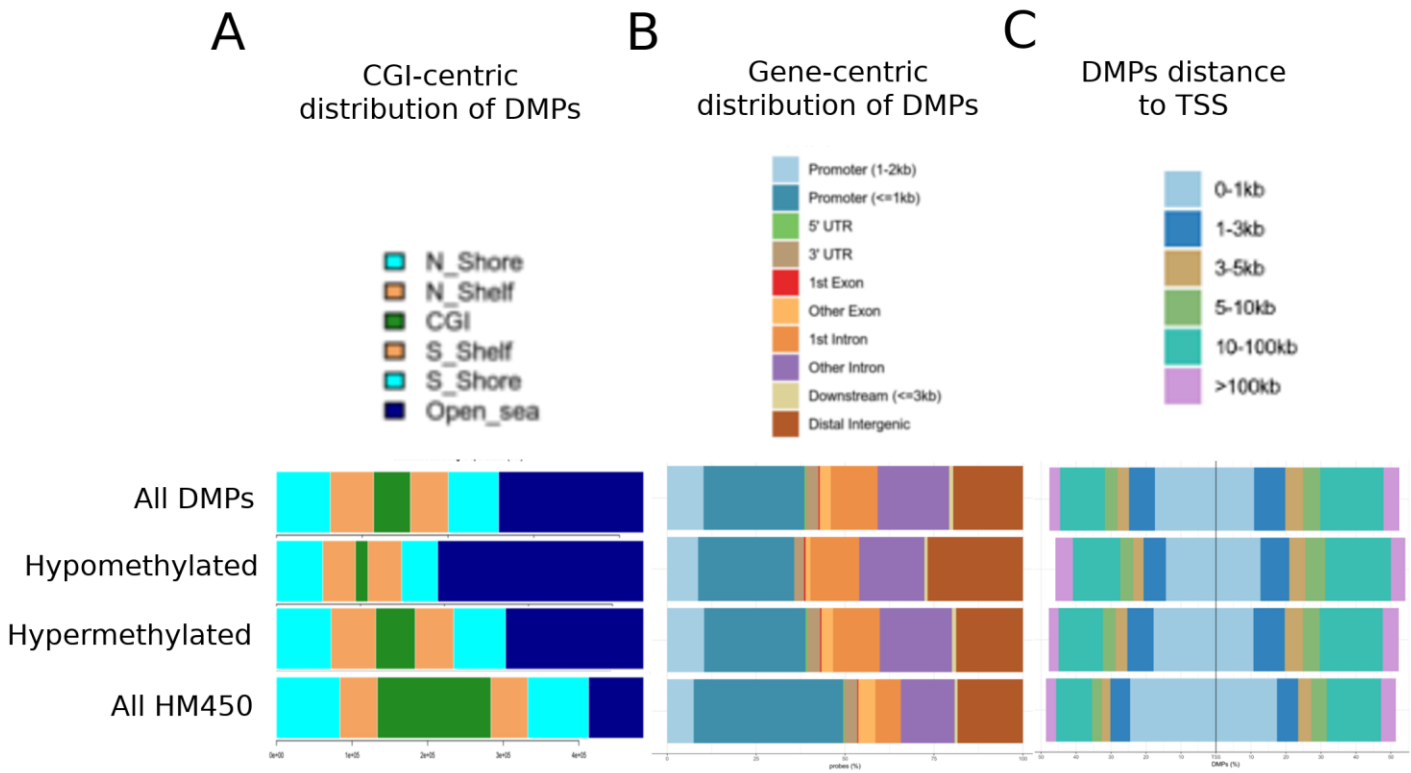
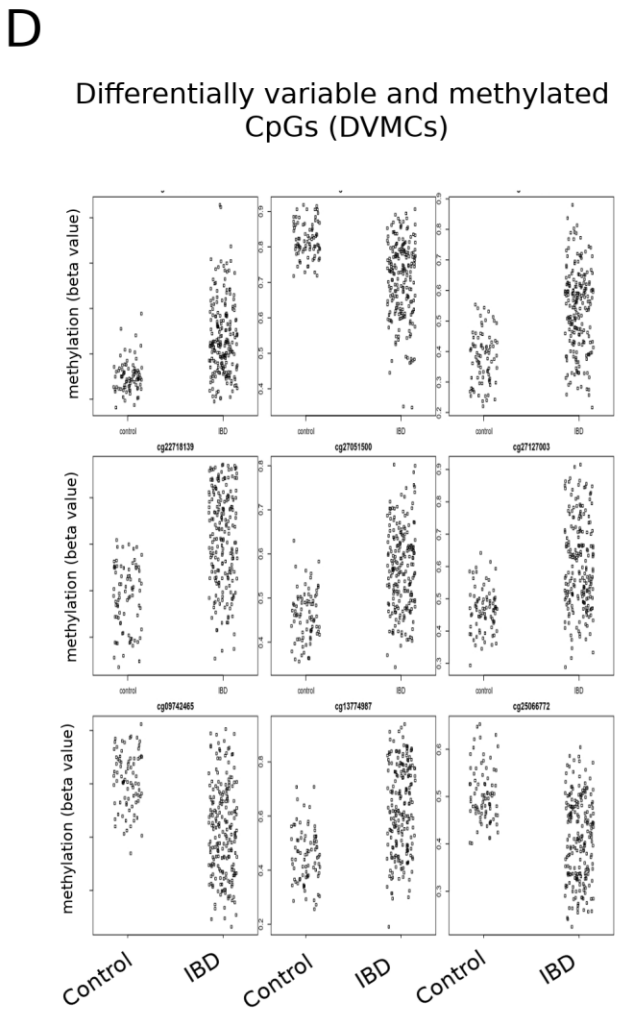
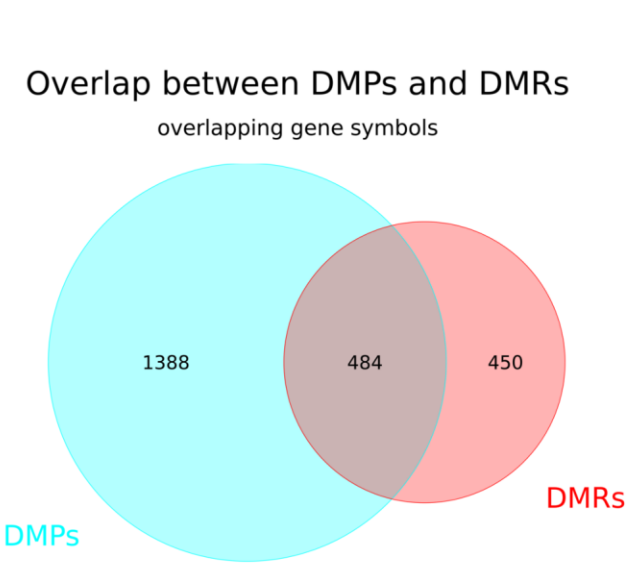

A
cg16465027 (PHACTR1)
cg07839457 (NLRC5)
cg16240683 (ZNF436-AS1)
cg19269426 (GGPS1)
cg22718139 (HMGCS2)
cg26974214 (LIPA)
Normalized Beta values
Control
IBD
B
CD
Control
UC
C
DMPs (delta_beta 20%)
IBD vs. Control
cg02806715 (HLA-DMA)
cg24129356 (HLA-DMA)
cg26974214 (LIPA)
cg01804934 (HLA-DPA1)
cg09321817 (HLA-DPA1)
cg06061086 (FOXP4)
cg23045908 (PDE4B)
cg07839457 (NLRC5)
cg16465027 (PHACTR1)
cg19269426 (GGPS1)
cg22718139 (HMGCS2)
cg16240683 (ZNF436-AS1)
Condition
control
IBD
Dataset
EPIC MTAB_5463
HM450 GSE105798
HM450 GSE32146
HM450 GSE42921
HM450 GSE81211
HM450 MTAB_5463
Localization
asc_col
ND
sig_col
ter_ile
Sex
F
M

Figure 1


A
CGI-centric distribution of DMPs
N_Shore
N_Shelf
CGI
S_Shelf
S_Shore
Open_sea
B
Gene-centric distribution of DMPs
Promoter (1-2kb)
Promoter (<=1kb)
5' UTR
3' UTR
1st Exon
Other Exon
1st Intron
Other Intron
Downstream (<=3kb)
Distal Intergenic
C
DMPs distance to TSS
0-1kb
1-3kb
3-5kb
5-10kb
10-100kb
>100kb
All DMPs
Hypomethylated
Hypermethylated
All HM450


D
Differentially variable and methylated CpGs (DVMCs)
methylation (beta value)
Control
IBD


E
Overlap between DMPs and DMRs
overlapping gene symbols
1388
484
450
DMPs
DMRs

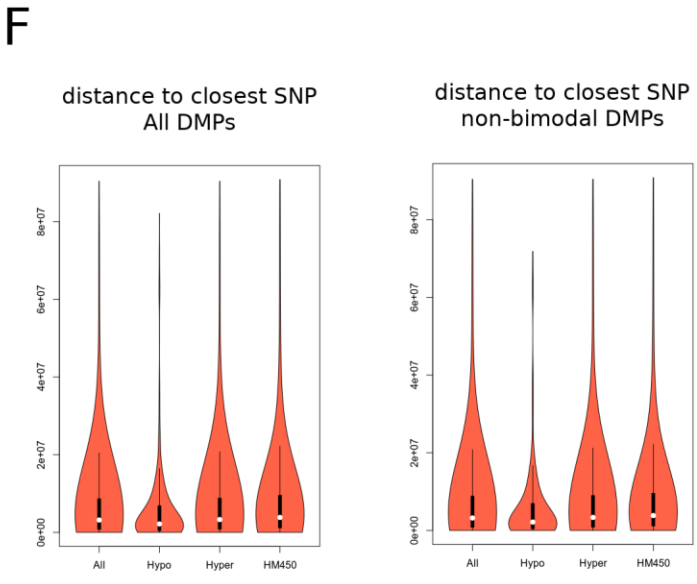
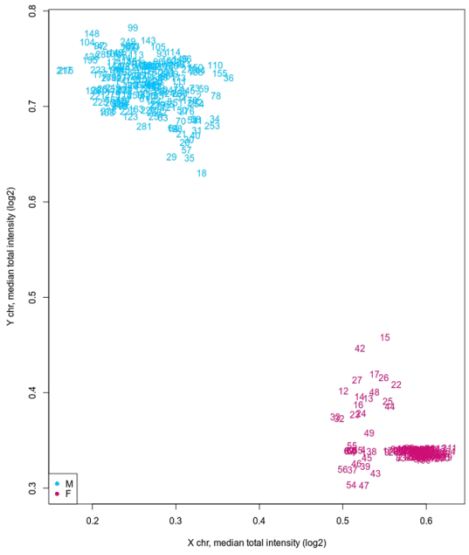
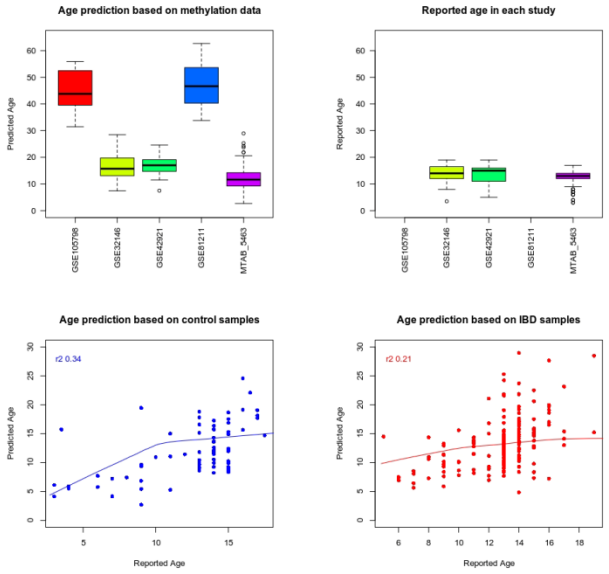

F
distance to closest SNP
All DMPs
distance to closest SNP
non-bimodal DMPs
All
Hypo
Hyper
HM450

Figure 2

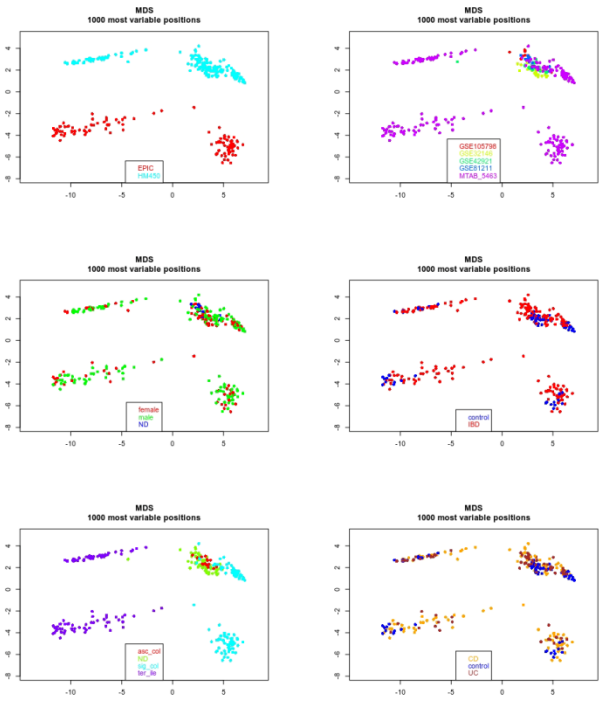
A



B



C



D

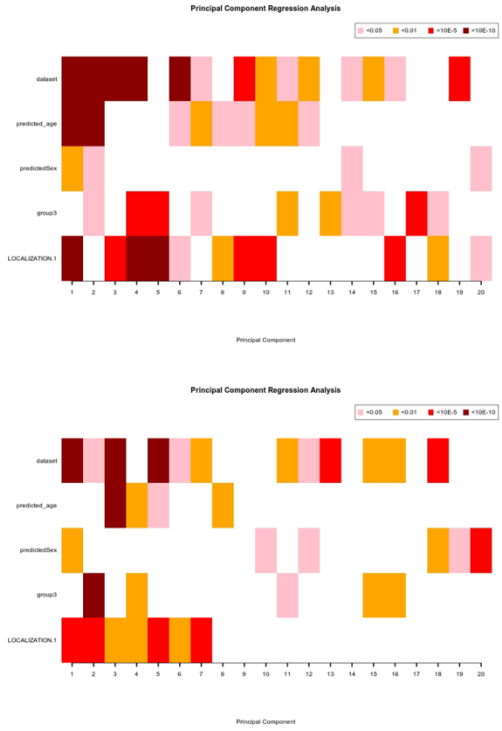
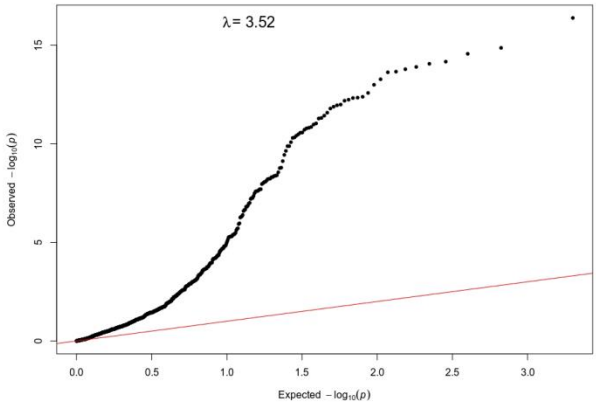
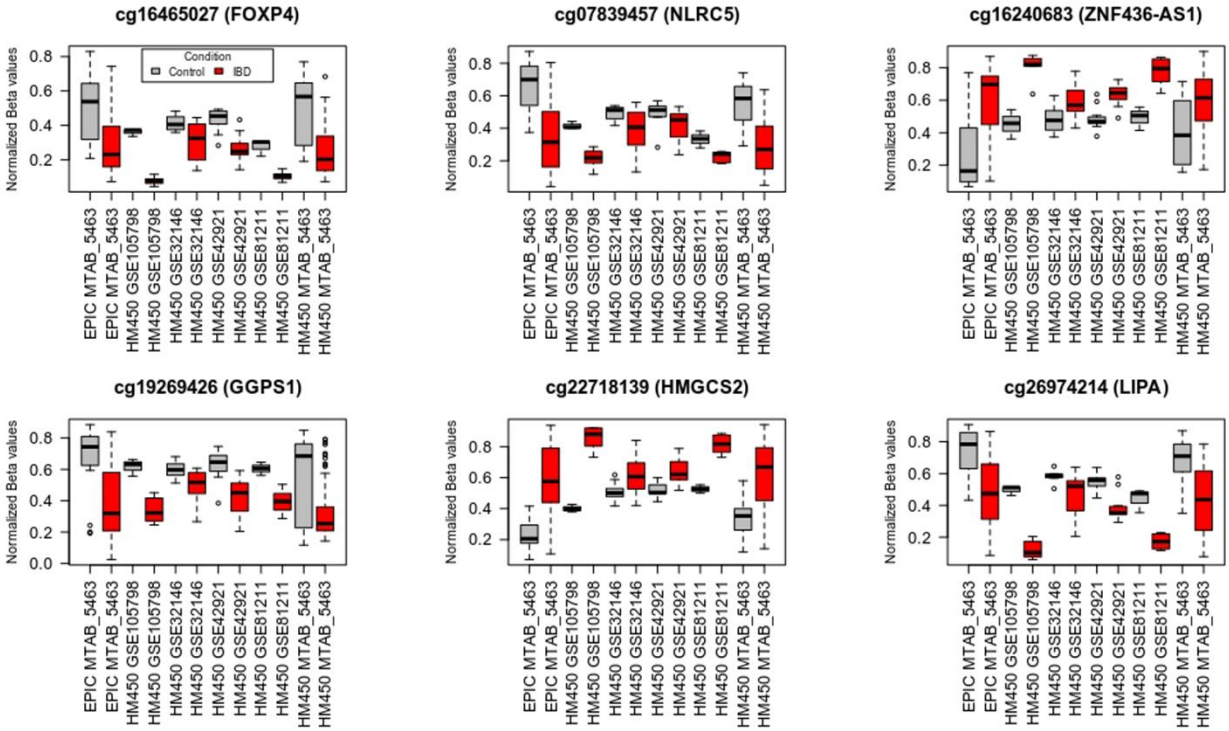


Figure S1

A



B



C

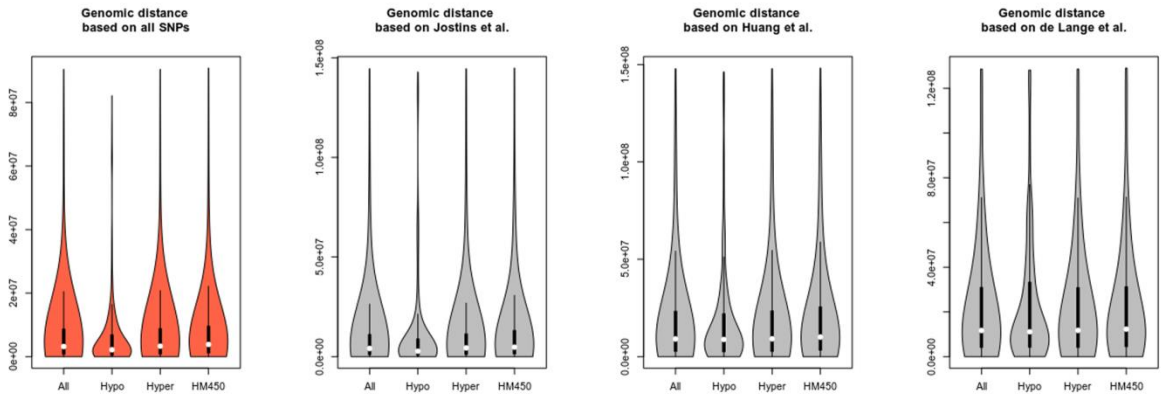


Figure S2